DOCUMENT RESUME

ED 113 693 CS 002 196

TITLE Development of the Brief Test of Literacy. National Center for Health Statistics, Series 2, No. 27.
INSTITUTION National Center for Health Statistics (DHEW), Rockville, Md.
PUB DATE Mar 68
NOTE 36p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Educational Assessment; Educational Research; *Evaluation Methods; Health; *Literacy; *Measurement Instruments; *Reading Skills; *Test Construction; Writing Skills
IDENTIFIERS *Literacy Tests

ABSTRACT

This report outlines the procedures involved in the development of a test of literacy suitable for use in screening large numbers of persons. The authors discuss the problems which were faced from the initiation of the project through the final assembly of the test materials, describing the difficulty of definition, the practical constraints on the administration, and the limitations of test design. On the basis of its use thus far, the resulting instrument, referred to as the "Brief Test of Literacy," appears to discriminate quickly and accurately between literate and illiterate persons. (Author/RB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED113693

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NATIONAL CENTER For HEALTH STATISTICS | Series 2 Number 27

VITAL and HEALTH STATISTICS

DATA EVALUATION AND METHODS RESEARCH

development of

the Brief Test of Literacy

A description of the development of a new test of literacy capable of providing information concerning the prevalence of illiteracy in large sample populations.

CS002196

Washington, D.C. March 1968

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
John W. Gardner
Secretary

Public Health Service
William H. Stewart
Surgeon General

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402 - Price 30 cents

NATIONAL CENTER FOR HEALTH STATISTICS

THEODORE D. WOOLSEY, *Director*

PHILIP S. LAWRENCE, Sc.D., *Associate Director*

OSWALD K. SAGEN, PH.D., *Assistant Director for Health Statistics Development*

WALT R. SIMMONS, M.A., *Assistant Director for Research and Scientific Development*

ALICE M. WATERHOUSE, M.D., *Medical Consultant*

JAMES E. KELLY, D.D.S., *Dental Advisor*

LOUIS R. STOLCIS, M.A., *Executive Officer*

DONALD GREEN, *Information Officer*

DIVISION OF HEALTH EXAMINATION STATISTICS

ARTHUR J. McDOWELL, *Director*

JAMES T. BAIRD, JR., *Chief, Analysis and Reports Branch*

HENRY W. MILLER, *Chief, Operations and Quality Control Branch*

PETER V. HAMILL, M.D., *Medical Advisor*

LAWRENCE E. VAN KIRK, D.D.S., *Dental Advisor*

LOIS R. CHATHAM, PH.D., *Psychological Advisor*

Public Health Service Publication No. 1000-Series 2-No. 27

Library of Congress Catalog Card Number 67-62374

FOREWORD

The Health Examination Survey, one of the major programs of the National Center for Health Statistics, collects, analyzes, and publishes the kinds of health-related data which can be obtained only through direct examinations, laboratory tests, and measurements. Much of the data collected pertains to prevalence levels of specific, medically defined diseases. Other data provide, for the population studied, distributions of a variety of physical, physiological, and psychological measurements. Reports in Series 1 and Series 11, described in the outline at the back of this publication, present the descriptions and some of the findings of the various programs already carried out.

In planning the third program of the series of Health Examination Surveys, consideration was given to including some measure of the extent of illiteracy in the population. That there is some relationship between various states of ill health and illiteracy has been recognized. It seemed desirable, therefore, to be able to investigate the relationships between some of the health findings and this measure. In addition, officials in other parts of the Department of Health, Education, and Welfare expressed interest in obtaining such data.

The usual procedure followed in planning programs of the Health Examination Survey is to utilize tests, procedures, and instruments already well established and generally accepted. In some instances, however, the special requirements of the survey along with the "state of the art" of measurement of the particular variable make this impossible. This is discussed in the present publication. In this instance, presented with such a problem, it was decided to enter into a contract with the Educational Testing Service to develop the required instrument. The results are presented in this report.

It is not surprising that the National Center for Health Statistics should sponsor such research. The Public Health Service is authorized under the National Health Survey Act (PL 652: 84th Congress) "to provide (1) for a continuing survey and special studies to secure . . . statistical information on the amount, distribution, and effects of illness and disability in the United States . . . and (2) for studying methods and survey techniques for securing such statistical information, with a view toward their continuing improvement."

The results of this study are being made available, not only to provide necessary information for evaluating later reports of findings in the Health Examination Survey programs, but also because of their more general interest. The report will call attention to the need for technically superior, yet brief, psychometric instruments, and it will inform interested persons and groups as to what has been done, in one instance, to meet this problem.

Arthur McDowell, Director
Division of Health Examination Statistics

SYMBOLS	
Data not available	---
Category not applicable	...
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

CONTENTS

	Page
Foreword	iii
Introduction	1
General Background	1
Establishing Test Specifications for Reading	2
Establishing Test Specifications for Writing	5
Pretests and Their Results	6
Construction of the Final Form	10
Screening Tryouts	12
Summary	12
References	12
Acknowledgments	13
Appendix I. Discussion of the Use of Phi Coefficients	14
Appendix II. Description of the Coefficient of Sentence Consistency	15
Appendix III. Answer Sheets for Reading and Writing Tests	16
Appendix IV. Instructions for Reading	18
Appendix V. Five Items Used in Writing Test	22
Appendix VI. Basic Skills Survey, Reading and Writing Manual for Examiners	23
Introduction	24
Administering the Reading Test	24
Scoring Information	26
Administering the Writing Test	27
Scoring Information	27

THIS REPORT outlines the procedures involved in the development of a test of literacy suitable for use in screening large numbers of persons.

In it the authors discuss the problems which were faced from the initiation of the project through the final assembly of the test materials, describing the difficulty of definition, the practical constraints on the administration, and the limitations of test design.

On the basis of its use thus far, the resulting instrument, which will be referred to as the Brief Test of Literacy, would appear to discriminate quickly and accurately between literate and illiterate persons. This report should provide valuable information to any prospective user of the test or to those who seek to develop their own instruments in this field.

DEVELOPMENT OF
THE BRIEF TEST OF LITERACY

Thomas F. Donlon and W. Miles McPeek, *Educational Testing Service*
Lois R. Chatham, *Division of Health Examination Statistics*

INTRODUCTION

The Brief Test of Literacy was developed to assess literacy in reading and in writing within the framework of a national health survey. As such it provides an instrument of marked utility, for no prior test intended for the direct assessment of literacy has been developed.

There are several reasons for the lack of any earlier development of a comparable instrument. In general, psychological testing has concentrated on the development of instruments which are appropriate for the measurement of individual differences, with a concomitant interest in the longer tests that are necessary to achieve high reliability. Only recently has there been any strong interest in instruments that are specifically intended to provide information concerning the educational attainment of groups. While instruments capable of such description will be developed with increasing frequency in the near future, the Brief Test of Literacy is one of the first of its type.

A second reason for the absence of a test of this kind is the concept of literacy. It is virtually impossible to achieve a satisfactory definition of literacy. It is even more difficult to attain an operational definition, and yet an operational definition is a virtual *sine qua non* for the development of a psychological test. The problem of definition is confounded by the varying demands of different cultures and subcultures and by cultural change through time. As a result, a person who is virtually illiterate by the standards of an advanced culture may well be able to meet the demands of his own less-developed civilization.

A third reason for the absence of an earlier test of this nature is that a large number of reading tests already exist. Many of these tests are intended to measure reading skill at approximately the level required. However, such existing tests can make a limited contribution to a survey of literacy because they are primarily designed either to evaluate children who are in the first years of school or to provide diagnostic information concerning the nature of reading problems, rather than to provide categorical assessment of literacy versus illiteracy.

For these reasons, the Brief Test of Literacy represents an initial development both in the general field of survey instruments and in the assessment of literacy.

GENERAL BACKGROUND

The Brief Test of Literacy was developed for the purpose of assessing literacy in reading and in writing within the framework of the National Health Survey whose mission is to study the incidence and prevalence of various health and health-related problems. Because of the nature of the survey, many different measures are obtained for each sample person; therefore, the amount of time allotted for the assessment of any one aspect of health is extremely limited.

As a result one of the primary constraints placed on the test was that it be so designed that literacy could be determined in a brief period of time. Toward this end a target time of 5 to 8 minutes was established.

In addition to the time restraint, the test had to be suitable for use with the general population of adolescents throughout the continental United States and, hopefully, with adults as well. Since the survey population excluded institutionalized persons, the test did not need to be designed to permit the rapid assessment of literacy in cases where the individual could not function in normal society because of extreme emotional disturbance or severe mental retardation.

A third constraint on the test was that it had to be so designed that the results could be interpreted in terms of the prevalence of literacy and of illiteracy. Accordingly, the fundamental measurement concept was that of a cutting score. Anyone above a designated score would be considered literate; those below it would be considered illiterate. Degrees of literacy would not be assessed.

ESTABLISHING TEST SPECIFICATIONS FOR READING

The initial step in the development of specifications consisted of a survey of the literature. This survey was disappointing. In spite of extensive work on the importance of literacy and on projects for its improvement in various nations, there were no reports on techniques for its direct assessment. In fact, as stated in the introduction, there is a general vagueness as to what constitutes literacy, with sundry definitions put forth by various writers. The most surprising finding was the absence of any general description of the assessment of literacy during World War II. There undoubtedly was extensive work in the area at that time: the military differentiated among low-level inductees, determining who should be given a basic education course, but there was nowhere a summary of the devices used. From a private communication with a government psychologist, it was learned that at present the Armed Forces use a general aptitude test to make these distinctions. This practice could not be followed by the survey, however, because of the obvious confounding of low mentality and of illiteracy.

While no specific techniques were uncovered in the literature search, a variety of definitions was found. In general, these fell into two classes, the functional and the normative. Functional definitions stressed an individual's adjustment to his culture. One was literate if he possessed a level of ability sufficient to permit him to function well in his society. Normative definitions stressed some typical educational attainment. Thus, one was literate if one read as well as the average child at the middle of the fourth grade in the United States, or at the end of the fifth year in Pakistan, et cetera.

The functional definition is inherently attractive, for illiteracy is a functional deficit. At the present time, however, there simply is no realistic basis on which to determine a functional level for a society as diverse as that of the United States; to attempt to describe the criteria for using such a definition would be a truly formidable task. The following quotation of a UNESCO definition[1] is an example of the difficulty.

> *A person is literate when he has acquired the essential knowledge and skills which enable him to engage in all those activities in which literacy is required for effective functioning in his group and community, and whose attainments in reading, writing, and arithmetic make it possible for him to continue to use these skills towards his own and the community's development and for active participation in the life of his country.*

One would be hard-pressed to translate these generalities into measurement specifics.

Therefore, in conjunction with the administration of the survey for which the test was to be developed, it was decided to estimate the incidence of illiteracy using a definition which is commonly held in the fields of education and health in this country, namely, "literacy is that level of achievement which is attained by the average child in the United States at the beginning of the fourth grade."[2]

With the establishment of a working definition, the development of the statistical specifications was begun. As stated earlier, the test was to be designed so that test scores could be assigned to one of two categories. The requirement built

in another specification—the use of a cutting-score technique. Given the working definition, the cutting score would ideally be such as to minimize the error in differentiating the top 50 percent of the national population of children entering fourth grade from the bottom 50 percent. The item statistics should be specified so as to achieve, then, this optimal cutting score.

The theory of the cutting score is quite complex. Major theoretical work in the area has been undertaken by Lord,[3] and there are fairly sophisticated techniques for developing such tests and locating the "cut." For various practical reasons, however, a more pragmatic approach was used in developing the Brief Test of Literacy. This pragmatic approach did retain one obvious feature of virtually all cutting-score work: the difficulty of the items was centered on a narrow band, rather than allowed to vary widely. This is in contrast to tests designed for differentiating among several levels of ability.

A practical limitation also arose in connection with the timing of the developmental work relative to the school year. The working definition of literacy was defined as achievement at the beginning of the fourth grade, but the developmental work had to be performed during the late winter months. If the scores made during winter months were to serve as estimates of the comparable difficulties which would be obtained using an entering fourth grade population, some adjustment in the observed item difficulties was needed. There was, however, no adequate empirical basis for determining this adjustment. After a review of available data on the growth of reading ability, it was decided that an average item difficulty level of 80-percent-pass at the time of pretesting would be a useful estimate of a difficulty of 50 to 60 percent for entering fourth graders. Accordingly, the specifications for item difficulty were set as follows: the items would show an average difficulty of 80-percent-pass and a range of difficulty from 65-percent-pass to 95-percent-pass.

The difficulty of the reading materials was specified to be approximately fourth-grade level. Deviations were permitted only in the direction of greater difficulty, because of the intended use of the materials with an older population and because the normal conception of reading difficulty is based partly on dimensions of reading beyond the kind of literal comprehension which was envisioned for this test. This limitation to literal comprehension is discussed later in the description of the type of questions asked. The conclusion was, however, that normal estimates of passage difficulty were likely to be overestimates, given the simplicity of the questions.

In the absence of any external criterion, item validity was limited to an index of internal consistency; phi coefficients[a] were specified as the indexes of item-test correlation. No specific mean value of these was established. Instead it was specified that the mean of the phi coefficients be maximized and that all items should show a phi coefficient significantly greater than chance.

The number of items in the test was also left unspecified. In a sense, there were incompatible goals for the proposed test in that test reliability had to achieve an acceptable level, while the time required for administration had to be minimized. A reliability between .70 and .80 was considered desirable for this survey work, and the ideal testing time was 5 minutes per person. At the beginning, the format of the test was uncertain. Clearly, there would be a presentation of material to be read, and there would be questions to determine comprehension, but the severe time constraints posed some difficulty in the development of test format. In any reading test there is usually an average ratio of the number of words which must be read for each question. This ratio must be large enough that a reasonable test of reading can be attained, and it must be small enough that testing time can be efficiently used. The problem posed in the test development work was the estimation of a workable value for this ratio.

Careful study led to the conclusion that the optimum format would consist of a brief passage of 40 to 50 words followed by two or three questions. Thus, another specification was established:

---

[a]The phi coefficient is a measure of the correlation between two variables when the variables are divided into quantitatively discrete groups and thus can be represented in a four-fold table. It is identical to the product moment correlation between two binomial variates. The phi coefficient is discussed in a number of statistical textbooks: for example, see Walker and Lev, *Statistical Inference*, New York, Henry Holt and Co., Inc., 1953.

the length of the passage to be read. The decision was also made to use three questions with each reading passage on the pretest. Ultimately, a decision would need to be made as to the use of two or three questions in the final form. This decision could be based both on the speed factor and on the patterns of losses of items due to defects uncovered in the pretesting.

Timing was a central concern. Reading proficiency has always consisted of a combination of two abilities: the ability to read rapidly and the ability to read accurately. Some reading tests attempt to provide diagnostic information as to the relative proficiency along these two dimensions. Generally the close correlation between the two measures, speed and accuracy (or comprehension), poses no real difficulty. However, at the level of skill required to make a judgment of literacy, less emphasis should be placed on speed as the source of variation among scores. Certainly, in a functional sense, speed of reading is important in achieving literacy; nevertheless, many poor readers must have time to allow the words to come into focus before they can establish meaning. It was obvious that, given the need for a 5-minute test, no power measure could be provided. Every effort was made, however, to reduce speed variance to a minimum.

One underlying consideration in establishing time specifications was not essentially psychometric, but it was such a powerful consideration with those working on the test that it deserves mention. "Illiteracy" is not a complimentary attribute, and although it is capable of specific redefinition in an operational sense—"an 'illiterate' is one who does poorly on our test"—the popular conception of illiteracy cannot be ignored. This popular conception undoubtedly stresses comprehension in reading far more than speed. In other words, to the extent to which it was possible, the test construction process limited speed variance to a level which seemed reasonable. The reading rates demanded by the test are not stringent in comparison with the everyday demands of our society.

Since a random sample of noninstitutionalized persons aged 12 through 17 living in the continental United States would be drawn in the survey, the typical sample subject should encounter no difficulty with the test. The poorer readers however, for whom there would exist a question of literacy, might have problems simply because of unfamiliarity with any testing situation. The multiple-choice format was specified for the reading test because of the efficiency it offered in response time and in scoring time. The use of a separate answer sheet, as opposed to a test booklet in which answers are recorded directly, posed certain problems. For example, a subject might fail to correctly align his answer sheet and test booklet, leading to invalid test scores. However, since the use of an answer sheet made it easier for the examiner to keep track of the subject's progress and to stop the examination when the cut-off score was achieved, the answer sheet method was adopted.

One concern remained. In a test of 5 minutes' duration, a subject of borderline intellectual ability who is not used to taking tests might, if left to himself, fail to divide his time properly. Thus he might spend too much time on one particularly difficult question and thereby score poorly on the whole test. Such personal characteristics are a cause of concern even in much longer tests. Because it was decided to avoid "speededness" in all of its forms, personal characteristics seemed even more likely to cause difficulty. To control for such variables, the test was made to consist of a number of separately timed units, monitored by the examiner to insure that the appropriate pace was maintained.

There were other reasons for developing a test of several parts. Foremost among these was the opportunity it would provide for shortening the total testing time for any subject who succeeded in passing the cutting score. Such a subject could complete the part on which he was working but would not need to attempt later parts. Another advantage would be derived in that an error in test administration during one of the parts need not require a complete retesting; rather, one additional section could be added to replace the defective one. The parts were specified to be separately timed units, consisting of a passage and two or three questions. At this point no decision was made concerning the amount of time which would be devoted to each passage; however, this was anticipated to be 60 or 90 seconds, depending on the outcome of the pretesting.

The scoring formula was specified as the total number of right answers minus one-fourth of the

number of wrong answers. While this is standard practice in multiple-choice testing, it was particularly indicated in this test, where the relatively few questions asked would make it possible to secure a substantial change in rank position merely by chance, if only the number of correct answers were used in the scoring.

Specifications regarding the content of the test were difficult to define. Perhaps the clearest specification was that the content had to be acceptable to adults and to adolescents, yet had to lend itself to pretesting on fourth graders. That is, materials from a storybook written for 10-year-olds would be inappropriate for adults. On the other hand, materials which would be pretested on a group of fourth graders could not contain language or topics inappropriate for children. In addition, materials had to be suitable for use with highly diversified populations. For example, the test had to be equally acceptable to boys and to girls, to persons with a science interest and to those with an art interest, to those who lived in the country and to those who lived in the city, to Negro and to white, and to rich and poor alike.

The anticipated use of the test on older populations led to the pretesting of a number of passages aimed at simulating the functional reading demands of adult life. These were in the form of want ads and brief instructions for operating equipment.

One important specification concerned the type of question which could be asked. In a reading test there is typically a variety of questions differentiated by the degrees of inference and judgment required to answer them correctly. Both inference and judgment play a role in reading ability, and each may be argued to be essential to literacy, in one of its meanings. These more complex aspects of reading would be excluded from the definition of literacy used in developing this test. Instead questions would be limited to straightforward comprehension. As a result all answers would be essentially restatements of information presented in the reading passage. While no defense of this decision may be necessary, it may be restated that any definition of literacy is an arbitrary dichotomization of what is fundamentally a continuum of varying reading ability from little or none to highly developed. Reading has dimensions, and it is possible to be more literate in one of these dimensions than in another. The most basic dimension in reading is straightforward comprehension, and the Brief Test of Literacy focused on this.

When the foregoing work had been completed, the test specifications for the reading test were virtually complete and the development of pretest materials was begun.

ESTABLISHING TEST SPECIFICATIONS FOR WRITING

Very early in the development of the writing test the decision was made to use the technique of having the subject write a few brief, simple sentences in response to dictation by an examiner. The writing test, because it called for a constructed response, required the development of a scoring technique which would be efficient for the examiner, consistent when used by varying scorers, and valid in its differentiation among subjects. A central problem in developing this scoring technique was that of spelling accuracy. If a person writes "Kum kwik wid the dokter," it is difficult to say he is illiterate. On the other hand, not all variations in orthography are so readily translated, and it is difficult to judge when a message has been conveyed and when it has not. Similar remarks pertain to handwriting legibility. It was specified that the subject's response could be either in printing or in cursive writing. Some highly literate persons produce a cursive script of formidable difficulty. How could such products be fairly evaluated?

It was decided that a two-dimensional approach, incorporating both a summation of the correctness of particular words and a global judgment of the sentence by the examiner would be used. As stated below, however, this specification was subsequently abandoned on the basis of pretest results.

While the specification of writing sentences as dictated was a practical decision, its central importance should not be overlooked. Literacy in writing is typically conceived as the ability to produce, rather than reproduce, a satisfactory message. Ideally, one would call for any sort of written message from the subject, allowing the subject to determine its content. The message then would be evaluated in some manner. Such

evaluations would be susceptible to variation, however. Furthermore this would lead to a variety of vocabulary samples since all subjects would not use the same words. Even worse, vocabulary content might well be selected by the subject to insure his success.

The total time required for the writing test was left unspecified. The time allotted for writing a given sentence was set at 1 minute, subject to modification following the results of the pretesting. Sentence length was to be approximately 10 words. Sentence topics were to stress practical situations, such as instructions.

The statistical specifications for the sentences could be more general, since the score variance would be spread over more categories than the simple right-wrong of the multiple-choice items used in reading. No precise difficulty measure was specified; an index of consistency with a total score and with the reading score was required but left unspecified until the nature of the scoring process was better defined.

Consideration was given to a format in which the subject would complete a brief document such as an application blank. This would be, in a sense, the analog of the "want ad" passages which were introduced into the reading test. This format was rejected because it would produce responses which were unique to the individual; what it offered in face validity for evaluating adult literacy, it would lose in comparability of subject performance.

PRETESTS AND THEIR RESULTS

In all, 25 reading passages and 75 questions were presented. The pretest population consisted of 180 fourth-grade students selected from public schools considered by the administrative officers of the school system to be about average in terms of national norms on ability tests. One minute was allowed for each passage and for its three questions. Observation of the first group confirmed the appropriateness of this timing. The responses were indicated by circling the answer in the test booklet directly, rather than by use of an answer sheet, because the mastery of an answer sheet is sometimes not complete among fourth-grade pupils and because the group administration procedure used in the pretest precluded the individual attention which could correct this.

Table 1. Grade level and number of words used in each pretest passage

Passage number	Grade level[1]	Number of words
1	4.1	44
2	4.8	44
3	4.2	49
4	4.9	57
≠5	6.3	33
6	5.9	56
7	5.7	52
8	5.0	35
9	4.8	55
≠10	6.5	39
11	5.6	54
12	4.8	47
13	4.1	36
14	5.7	52
≠15	8.6	53
16	4.4	39
17	4.9	40
18	4.5	43
19	4.5	35
≠20	5.7	64
21	4.9	57
≠22	6.1	51
≠23	7.2	61
24	4.7	48
≠25	7.3	60
Average	5.4	48.2
Range	4.1-8.6	33-64

Grade level frequency distribution

8.0-8.9	1
7.0-7.9	2
6.0-6.9	3
5.0-5.9	6
4.0-4.9	13

[1]Determined by Lorge formula.

≠"Adult" content (i.e., material from want ads or instruction manuals).

The success of the pretest demanded that the judgments of difficulty be quite accurate. As a check on these judgments, the index of reading difficulty proposed by Lorge[4] was computed for each passage. This index takes into account such factors as the length of the sentences and the number of "uncommon" words (defined as any words not included in the Dale-Chall list of basic words).

Data concerning this index and passage length are presented in table 1. This table shows that

the average Lorge index was 5.4—that is, it corresponded in difficulty to the level of material with which the average pupil can cope in about the fourth month of the fifth grade. This figure was quite a bit higher than either the grade-level index of the pretest population, which was 4.5, or that of the theoretical reference population, which was 4.0. It was felt that this was justified because the group for which the materials were being developed would be over 11 years of age and therefore, theoretically, beyond fourth-grade placement. Furthermore, the questions in the literacy test would be limited to assessment of comprehension whereas the Lorge assessment was based on a complex of skills.

As stated in the discussions of the specifications, there was an attempt to develop materials with a higher "face-validity" for adults, as illustrated by items from want ads or instruction manuals which accompany appliances or equipment. In spite of efforts toward reducing the difficulty of this type of material, it constituted the seven most difficult passages in terms of the Lorge index, as indicated by the high values associated with the passages in table 1 which are marked with a dagger (†). Their possible value in securing subject acceptance was sufficiently great to warrant pretesting.

In addition to the 25 reading passages, 10 simple sentences were read aloud, with instructions to write them in the space provided. In general, there was more difficulty with the pretesting than had been anticipated, for writing in response to dictation is not a routine school activity at this level. Fortunately, the true simplicity of the task made it possible to elicit adequate responses with a minimum amount of assistance from proctors.

There were three related statistics used in the evaluation of the reading pretest results. First, for each question there was computed a phi coefficient measuring its consistency with the total formula score on the entire 75 questions for the whole group. Second, for each question there was computed a phi coefficient measuring its consistency with the total formula score for the bottom 40 percent of the total group. Finally, for each passage the sum of the phi coefficients of its questions, as determined on the bottom 40 percent, was computed. These statistical results are presented in table 2, together with information concerning the level of difficulty of the material (in terms of the percentage passing). As described in the footnote to this table, the phi coefficient for the total group is referred to as "phi 20-80," and that for the bottom 40 percent as "phi 50-50," reflecting the point at which the groups were divided. This point is, of course, actually the same in both cases, for the 20th percentile in the total group is the 50th percentile in the lowest 40 percent. Two different phi coefficients were required to insure effective differentiation of questions in the region of greatest interest. Appendix I presents a more extended discussion of this.

As indicated in table 2, the pretesting was generally successful. Of the 25 passages, 12 secured a cumulative sum of phi 50-50 which exceeded 100. Among the 36 questions which pertained to these passages, only 4 had related phi coefficients which failed to attain statistical significance at the .01 level of confidence (phi equal to or greater than .31), and 29 questions had coefficients significant at the .001 level (phi equal to or greater than .39).

The passages with "adult" content were unsuccessful, with the exception of passage number 22, largely because these passages were too difficult to provide differentiation among the bottom 40 percent. All of the 10 most difficult questions were associated with these materials. The general success of the difficulty estimation is indicated by the average difficulty of the questions which were not "adult" content. For these 18 passages, the average question was passed by 77 percent of the total group, which was very near the specified value of .80. One other point became clear in the pretesting. The third question was typically not much affected by "drop-out," the usual indication of "speededness." Accordingly, the use of three questions with each reading passage could be continued in the final form.

The writing pretest generally sustained the appropriateness of the 1-minute time limit. The assessment of the consistency between success on a given sentence and success on a total score for writing (or for reading) was not easy, as scoring procedures for the sentences had not been developed. Rough approximations were secured by scoring the sentences word by word, the test of a word being the judgment that it was legible and that its meaning was conveyed in

Table 2. Difficulty and validity indexes

Passage and item	Total group		Next-to-lowest fifth	Lowest fifth	Two lowest fifths	Cumulative sum of phi
1	Percent passing	Phi 20-80[1]	Percent passing		Phi 50-50[2]	
1	83	66	81	33	49	...
2	63	65	36	-	47	96
3	59	49	25	11	18	114
2						
4	90	53	92	58	39	...
5	78	47	61	14	49	88
6	81	47	72	16	56	144
3						
7	93	29	92	78	20	...
8	75	38	69	42	27	47
9	82	62	83	33	51	98
4						
10	88	45	83	58	27	...
11	82	49	75	44	32	59
12	69	45	42	28	15	74
#5						
13	85	45	86	58	31	...
14	45	34	25	11	18	49
15	34	24	22	11	15	64
6						
16	80	46	72	44	28	...
17	76	63	69	22	47	75
18	63	30	53	33	20	95
7						
19	79	49	64	39	25	...
20	78	50	64	36	28	53
21	64	61	39	6	40	93
8						
22	84	36	75	58	18	...
23	83	43	72	64	9	27
24	65	69	39	-	49	76
9						
25	86	52	89	50	42	...
26	74	76	75	8	68	110
27	63	53	36	11	30	140
#10						
28	63	25	44	39	5	...
29	76	66	81	19	62	67
30	31	21	6	11	-9	58
11						
31	84	54	86	44	44	...
32	83	67	86	33	54	98
33	78	67	72	22	50	148
12						
34	78	53	72	33	39	...
35	75	68	78	17	61	100
36	59	49	42	11	35	135
13						
37	89	53	92	56	41	...
38	83	56	81	42	40	81
39	82	77	86	22	64	145

Table 2. Difficulty and validity indexes

Passage and item	Total group		Next-to-lowest fifth	Lowest fifth	Two lowest fifths	Cumulative sum of phi
14	Percent passing	Phi 20-80[1]	Percent passing		Phi 50-50[2]	
40	87	56	89	50	42	...
41	64	52	36	14	25	67
42	72	65	58	14	46	113
†15						
43	70	59	61	17	45	...
44	57	49	31	8	29	74
45	21	12	6	11	-9	65
16						
46	80	48	89	42	49	...
47	86	60	97	44	58	107
48	80	72	83	22	61	168
17						
49	87	64	97	44	58	...
50	75	48	69	19	50	108
51	82	42	64	25	39	147
18						
52	89	66	97	47	56	...
53	75	43	67	39	28	84
54	76	66	78	19	59	143
19						
55	86	67	92	50	46	...
56	86	68	97	39	62	108
57	78	63	75	25	50	158
†20						
58	26	17	17	11	9	...
59	51	18	39	33	6	15
60	22	20	11	6	9	24
21						
61	74	50	61	31	30	...
62	66	52	53	17	38	68
63	66	46	44	22	23	91
†22						
64	81	50	89	42	49	...
65	44	39	33	6	34	83
66	53	42	42	11	35	118
†23						
67	53	28	33	25	9	...
68	52	33	25	19	7	16
69	45	29	17	16	1	17
24						
70	88	41	83	61	25	...
71	78	61	69	28	41	66
72	64	46	33	19	16	82
†25						
73	41	25	19	17	3	...
74	59	37	33	22	12	15
75	28	13	8	17	-14	1

[1] A phi coefficient based on splitting the total group into the top 80 percent and the bottom 20 percent.

[2] A phi coefficient based on splitting the bottom 40 percent into upper and lower halves.

† "Adult" content (i.e., material from want ads or instruction manuals).

Table 3. Mean score and range of scores, by fifths

Fifth	Mean score[1]	Range of scores
Highest scoring fifth--	20.47	18-21
Next-to-highest fifth--	19.66	17-21
Middle fifth----------	17.47	12-21
Next-to-lowest fifth---	14.61	8-18
Lowest scoring fifth---	2.92	-4-11

[1]Mean score computed as follows: Total number of correct answers minus one-fourth the number of wrong answers.

context. The distributions of these scores were then compared for the two lowest fifths, using a rough "consistency measure" which counted the number of times that those in the next-to-lowest fifth on total score were better on the particular sentence than those in the lowest fifth, and vice versa. The greater this "distance measure," the more the agreement between the score for each item and the total score. Because the sentences were of unequal length, however, they could not be readily compared. The labor of developing the complex statistical information which would provide a comparison was not justified. Virtually every sentence demonstrated a marked consistency with the total score; final selection was, in general, based on other factors. A brief description of the consistency measure is provided in Appendix II.

CONSTRUCTION OF THE FINAL FORM

On conclusion of the pretesting, the final phase of test specification and construction was undertaken. Of the 12 most successful passages, one passage (number 14) was eliminated because its cumulative phi depended greatly on the last question, raising the danger of "speededness." Then the pretest data were examined in order to determine an optimal number of passages for the final form. This number was approximately seven, or 21 questions. Accordingly, 7 passages were selected from the 11 possibilities. In this selection both item statistics and content were considered. Thus, pretest passage number 22 was preferred over passages with better statistics because of its "adult" content. The seven passages selected are the first seven shown in Appendix IV.

The total score characteristics of the seven-passage, 21-item test were examined. Table 3 presents the mean score using the formula, total number of correct answers minus one-fourth the number of wrong answers (R - ¼W), on the 21 items for the ability groups defined by pretest items and the score range observed in each group.

As shown, the test provides the greatest differentiation between the two lowest fifths and virtually none between the two top fifths. This is, of course, the desired characteristic. An additional investigation of the separation between the two lowest fifths is provided by table 4, which shows the score distribution for both.

The data in table 4 were the basis for the final decision concerning the location of the cutting score, which was set at 10.75 or greater. That is, persons scoring 10.50 would be classed "illiterate," and those scoring 10.75 would be

Table 4. Formula score frequency distributions (R-1/4W) for the two lowest total-score fifths

Score	Lowest fifth	Next-to-lowest fifth
-4----------	3	-
-3----------	1	-
-2----------	2	-
-1----------	3	-
0----------	4	-
1----------	3	-
2----------	4	-
3----------	1	-
4----------	1	-
5----------	2	-
6----------	-	-
7----------	4	-
8----------	3	2
9----------	4	-
10----------	-	-
11----------	1	4
12----------	-	1
13----------	-	1
14----------	-	3
15----------	-	11
16----------	-	7
17----------	-	4
18----------	-	3

Table 5. Mean frequencies for and differences between the two lowest fifths, by response category

Response category	Next-to-lowest fifth	Lowest fifth	Column 1 minus column 2
	Mean frequency		
Total--	21.00	21.00	...
Right--------	15.61	5.92	+9.69
Wrong--------	3.39	11.56	-8.17
Omitted------	-	0.19	-0.19
Not reached--	2.00	3.33	-1.33
Mean R-1/4W--	14.61	2.92	...

classed "literate" within the meaning of the working definition.

Table 5 presents a comparison of the two lowest fifths with respect to the average number of responses which fall into four basic categories: right, wrong, omitted, and not reached. Both "omitted" and "not reached" are blanks, with no response indicated on the answer sheet. An "omit" is a blank which is followed (not necessarily immediately) by a response to a subsequent question. An item is "not reached" if it is left blank at the end of a series of responses. "Not reached" responses are used to indicate "speededness" in a test; "omit" responses are generally considered to indicate ample time for a response but a failure to perceive the correct response. There is always ambiguity about the two categories: an "omit" may not have been read, due to pressure of time; a "not reached" item may have been considered. Nevertheless, the distinction offers some assistance in the quantitative assessment of speed.

As shown in table 5, there is a negligible amount of "speededness" in the test. The difference in score means between the two groups is 11.69; of this, only 1.33 is attributable to the difference in "not reached" items, and then only if the lowest fifth can be assumed to have perfect success on these items. In general, then, "speededness" is a very small factor in the test. Further, the small number of "omits" indicates that the items are not skipped as the test is worked through. Apparently the salient characteristics of the items are such as to encourage responding.

The reliability of the 21-item test was estimated to be .91 by a technique suggested by Raju and Guttman.[5] This estimate indicates an excellent reliability for the survey work for which the test is intended. Additional features of the test which heightened its utility for the survey were the use of the cutting score for securing briefer records and the availability of substitute passages for "repairing" a record damaged by the faulty administration of one of the passages.

The final development of the writing test was broadly similar to that of the reading test. A five-sentence test, totaling 47 words, with 1 minute per sentence was developed (see Appendix V). The five sentences were selected for appropriate consistency with a total-score criterion, for variety of content and vocabulary, and for sentence length. Once a scoring technique was developed, a cutting score was determined. This is between 27 and 28 (fractional scores are not possible): a person scoring 27 is classed "illiterate"; a person scoring 28 is "literate." This cutting score is estimated to divide subjects at grade level 4.0 into two equal groups on the basis of the data on the sample of subjects at grade level 4.5.

The principal labor concerning the writing test was the devising of a reliable scoring procedure. Initial attempts to develop a scheme which relied on judgment for accepting or rejecting homophone approximations to standard orthography ("dokter," "tumorow") proved unworkable. Even a group of staff members accustomed to working together on verbal items could not secure a sufficiently high degree of consistency. After much experimentation, it was decided to maximize the reliability of the scores by creating a scoring system which assigned a score based principally on errors of misspelling, of word inversion, and of word redundancy. This technique is described in the examiner's manual (Appendix VI). It has satisfactory correlation with the various subjective and judgmental approaches; what it loses in occasional instances by overpenalizing spelling errors, it gains in other cases by permitting different raters to score complex sentences in a similar manner.

Table 6. Length of time per test unit, based on performance of 12 students identified as poor readers

Passage and sentence	Average time in seconds
Passage	
1	47.5
2	53.3
3	43.1
4	50.4
5	42.5
6	49.3
7	50.3
8	47.4
9	46.3
10	46.0
11	47.3
Sentence	
1	27.3
2	34.8
3	33.3
4	33.9
5	38.4

SCREENING TRYOUTS

The construction of the final form was followed by screening tryouts in which the new instrument was administered in a person-to-person situation to 24 students aged 14 through 17 who had been identified by reading teachers as having reading difficulty. The purposes of this tryout were to assure that workable administration procedures were developed and that passage content was equally acceptable at the older age range, and to check on "speededness."

These trials were very successful. While no formal validity estimates were provided by the teachers, there was informal evidence in that the three persons who would be judged "illiterate" by the test were in fact so judged by the school. Expectations regarding the time element were confirmed. Even in this population, there was a considerable shortening of the time required as soon as any appreciable literacy was found. No use was made of the cutting score, since all passages needed to be screened for content acceptability, but the general practicality of the procedure was demonstrated.

An answer sheet enabling all responses to be recorded, both for reading and for writing, had been devised (see Appendix III).

Table 6 presents the average time required for each passage and for each sentence as observed by one examiner in screening tryouts. The given averages are based on only 12 cases, but the consistency of the results across passages and sentences lends credence to their reliability.

These average times demonstrate that while the total working time for all tasks can be as much as 12 minutes, this will not often be the case. The Brief Test of Literacy is indeed "brief."

SUMMARY

This detailed account of the developmental procedures has concentrated on description rather than on critical evaluation. Many of the steps involved were based on assumption or professional judgment, the adequacy of these being crucial to the success of the enterprise. Similarly, whereever statistical data were the basis for decision, the size of the sample from which they were drawn was a practical maximum rather than a theoretical optimum.

Nevertheless, the general consistency of the results and their coherence suggests that the developmental procedures were highly successful. It is expected that, following the establishment of norms and the completion of validation studies, the Brief Test of Literacy will provide a useful instrument for survey purposes.

REFERENCES

[1] "World Campaign for Universal Literacy." Document submitted by UNESCO in response to a request of the United Nations General Assembly at its Sixteenth Session. May 1963. p. 39.

[2] English, H. B., and English, A. C.: *A Comprehensive Dictionary of Psychological and Psychoanalytical Terms*. New York. David McKay Co., Inc., 1958.

[3] Lord, F. M.: Cutting scores and errors of measurement. *Psychometrika* 27:19-30, 1962.

[4] Lorge, I.: *The Lorge Formula for Estimating Difficulty of Reading Materials*. New York. Bureau of Publications, Teachers College, Columbia University, 1959.

[5] Raju, N. S., and Guttman, L.: A new working formula for the split half reliability model. *Educ. & Psychol. Measur.* 25(4): 963-967, 1965.

Acknowledgments

The development of the test materials and the conduct of the pretesting tryouts were facilitated within Educational Testing Service by the cooperation of Miss Susan Humphrey, Mrs. Helen Spiro, and Mrs. Sara Hufham. Further, the project could not have been completed without the cooperation of the public schools of the city of Trenton, N.J., through Dr. Sarah Christie, Assistant Supervisor of Schools, and the public schools of Princeton, N.J., through Mr. Thomas Seraydarian, Director of Guidance.

———o o o———

APPENDIX I

DISCUSSION OF THE USE OF PHI COEFFICIENTS

The need for two phi coefficients, as presented in table 1, may be most quickly demonstrated by the following contrived examples of contingency tables. In each case the entries in cells and margins are percentages of the *total group*.

The following question would show a sizable phi coefficient of consistency between item and test:

	TEST		
ITEM	10	70	80
	10	10	20
	20	80	100

Suppose, however, that the performance of the top 80 percent, in which seven-eighths or 87.5 percent were successful, was examined more closely as a 2 x 5 table in which each fifth of the total group is presented separately:

	TEST					
ITEM	10	10	20	20	20	80
	10	10	-	-	-	20
	20	20	20	20	20	100

Note that the item actually differentiates most markedly between the bottom 40 percent and the top 60 percent. In fact, phi 50-50 on the bottom 40 percent would be zero, correctly indicating that this item should not be chosen in spite of the value of phi 20-80.

On the other hand, there are anomalies in items, and it is the function of item analysis to guard against them. An item might show the following table for the bottom 40 percent, which would yield a sizable phi:

	TEST		
ITEM	5	15	20
	15	5	20
	20	20	40

Information on the top 60 percent, however, might lead to a completed 2 x 5 table of

	TEST					
ITEM	5	15	5	10	20	55
	15	5	15	10	-	45
	20	20	20	20	20	100

indicating that item ambiguity or some other peculiarity was distorting the normal pattern of increasing item success with increasing ability. For the 2 x 5 table, phi 20-80 would be computed from

	TEST		
ITEM	5	50	55
	15	30	45
	20	80	100

which would be lower than phi 50-50, signaling the distorted pattern.

The foregoing cases are necessarily preselected and dramatic. Nevertheless, the practice of using two coefficients of this type in the development of a cutting score instrument has much to recommend it.

APPENDIX II

DESCRIPTION OF THE COEFFICIENT OF SENTENCE CONSISTENCY

An example of the consistency measure used in evaluating the sentences is presented below. In general, a given sentence is consistent with the total score if those in the more able group score higher than those in the poorer group. The consistency measure totals the number of times that an individual in the superior group scores higher than an individual in the less able group; from this total is subtracted the number of times that individuals in the less able group surpass individuals among the superior group.

Suppose that a given six-word sentence yielded the following distributions:

Sentence score	Total score	
	Lowest fifth	Next-to-lowest fifth
6	-	5
5	-	5
4	5	10
3	5	10
2	10	5
1	10	5
0	10	-
	40	40

The consistency measure would be computed as follows:

(1) Score	(2) Number of superior group in category	(3) Number of less able group they surpass	(4) Number of less able group who surpass them	(5) (2)x[(3)-(4)]
6	5	40	-	200
5	5	40	-	200
4	10	35	-	350
3	10	30	5	250
2	5	20	10	50
1	5	10	20	-50
			Index	1,000

The chance expectation of this index is zero, negative values indicate an inverse relationship, et cetera. It is similar to several such indexes proposed in the psychometric literature.

———o o o———

APPENDIX III

ANSWER SHEETS FOR READING AND WRITING TESTS

SAMPLE PAGE					
Question Number	Answer Choice				
01	A	B	C	D	E
02	A	B	C	D	E
03	A	B	C	D	E

Name _____

Question Number	Answer Choice					Question Number	Answer Choice					Question Number	Answer Choice				
1	A	B	C	D	E	10	A	B	C	D	E	22	A	B	C	D	E
2	A	B	C	D	E	11	A	B	C	D	E	23	A	B	C	D	E
3	A	B	C	D	E	12	A	B	C	D	E	24	A	B	C	D	E
4	A	B	C	D	E	13	A	B	C	D	E	25	A	B	C	D	E
5	A	B	C	D	E	14	A	B	C	D	E	26	A	B	C	D	E
6	A	B	C	D	E	15	A	B	C	D	E	27	A	B	C	D	E
7	A	B	C	D	E	16	A	B	C	D	E	28	A	B	C	D	E
8	A	B	C	D	E	17	A	B	C	D	E	29	A	B	C	D	E
9	A	B	C	D	E	18	A	B	C	D	E	30	A	B	C	D	E
						19	A	B	C	D	E	31	A	B	C	D	E
						20	A	B	C	D	E	32	A	B	C	D	E
						21	A	B	C	D	E	33	A	B	C	D	E

ANSWER SHEET FOR WRITING TEST

Name

1.

2.

3.

4.

5.

APPENDIX IV

INSTRUCTIONS FOR READING

ON EACH PAGE IN THIS BOOKLET THERE IS A SHORT PARAGRAPH WHICH IS FOLLOWED BY THREE QUESTIONS. BELOW EACH QUESTION ARE FIVE STATEMENTS, ONLY ONE OF WHICH MAKES A GOOD AND SENSIBLE ANSWER. YOU SHOULD FIND THIS STATEMENT, AND MARK YOUR ANSWER BY CIRCLING THE LETTER ON THE ANSWER SHEET WHICH CORRESPONDS TO THE STATEMENT YOU SELECT.

YOU MUST WORK AS QUICKLY AS YOU CAN, FOR YOU WILL BE ALLOWED ONLY ONE MINUTE TO WORK ON EACH PARAGRAPH. BECAUSE THE TIME IS SO SHORT, YOU MAY NOT FINISH ALL OF THE QUESTIONS. IF YOU DO FINISH A PAGE BEFORE THE TIME IS UP, TELL ME AND YOU WILL BE ALLOWED TO GO ON TO THE NEXT PAGE.

Department of Health, Education, and Welfare
Public Health Service
National Center for Health Statistics

Reprinted with Permission
of
Educational Testing Service
Princeton, N.J. Berkeley, Calif.
© Copyright 1966
All rights reserved

SAMPLE PAGE

It was a beautiful gift, wrapped with bright red paper and tied with silver string. It was small, but very heavy. No one knew who had brought it, but it had Mr. Jones' name on top. Mr. Jones just smiled and said, "I'll open it when I get home."

01. Whose name was on the top of the gift?

(A) Mr. Jones
(B) Mr. Pike
(C) Willy
(D) The postman
(E) No one knew

02. In what color paper was the gift wrapped?

(A) Red
(B) Silver
(C) Green
(D) Orange
(E) Yellow

03. Where was the gift going to be opened?

(A) Where it was found
(B) At the police station
(C) In the car
(D) At the office
(E) At home

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

- 0 -

It was spring. The young boy breathed the warm air, threw off his shoes, and began to run. His arms swung. His feet hit sharply and evenly against the ground. At last, he felt free.

1. What time of year was it?

(A) Summer
(B) Fall
(C) Spring
(D) December
(E) July

2. What was the young boy doing?

(A) Running
(B) Jumping
(C) Going to sleep
(D) Driving a car
(E) Fighting

3. How did he feel?

(A) Hot
(B) Free
(C) Angry
(D) Cold
(E) Unhappy

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

- 1 -

There were footsteps and a knock at the door. Everyone inside stood up quickly. The only sound was that of the pot boiling on the stove. There was another knock. No one moved. The footsteps on the other side of the door could be heard moving away.

4. The people inside the room

(A) Hid behind the stove
(B) Stood up quickly
(C) Ran to the door
(D) Laughed out loud
(E) Began to cry

5. What was the only sound in the room?

(A) People talking
(B) Birds singing
(C) A pot boiling
(D) A dog barking
(E) A man shouting

6. The person who knocked at the door finally

(A) Walked into the room
(B) Sat down outside the door
(C) Shouted for help
(D) Walked away
(E) Broke down the door

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

- 2 -

Helen liked going to the movies. Sometimes she went four times a week. Everyone said she was crazy. Why did she always want to go out and spend money, they said, when she could stay home and watch television?

7. What did Helen like to do?

(A) She liked to eat
(B) She liked to swim
(C) She liked to watch baseball
(D) She liked to watch movies
(E) She liked to watch wrestling matches

8. What did people think about her?

(A) They thought she was crazy
(B) They thought she was very smart
(C) They thought she was very nice
(D) They thought she was ugly
(E) They thought she was very old

9. What did people think she should do?

(A) Write a book
(B) Watch television
(C) Go on a diet
(D) Dye her hair
(E) Stop talking so much

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

- 3 -

You could smell the fish market long before you could see it. As you came closer you could hear merchants calling out about fresh catches or housewives arguing about prices. Soon you could see the market itself, brightly lit and colorful. You could see fishing boats coming in, their decks covered with silver-grey fish.

10. What kind of a market is described above?

(A) A vegetable market
(B) A meat market
(C) A fish market
(D) A flower market
(E) A fruit market

11. What could you see coming in?

(A) Tug boats
(B) Rowboats
(C) Passenger boats
(D) Fishing boats
(E) Sailboats

12. What covered the decks of the boats?

(A) Rope
(B) People
(C) Cars
(D) Boxes
(E) Fish

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

Bill settled down sleepily into the seat at the back of the bus. All he wanted to do was to sleep until it was time to get off. But the noise of a nearby radio and the voices of the passengers kept him awake. Without thinking, Bill stood up and shouted, "Shut up, everybody!"

13. In what was Bill riding?

(A) A boat
(B) A car
(C) A plane
(D) A taxi
(E) A bus

14. What did Bill want to do as he rode?

(A) Sleep
(B) Eat
(C) Drink
(D) Talk
(E) Read

15. What did he shout?

(A) "Help!"
(B) "This is my stop!"
(C) "Shut up, everybody!"
(D) "There's a fire!"
(E) "We're going to crash!"

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

Tiger is a large, yellow cat. At night he prowls outside and is very fierce. When he hears a noise, he lowers his head and walks with stiff legs. All the other cats are afraid to come into his yard.

16. When does Tiger prowl?

(A) At dawn
(B) At dinnertime
(C) In the afternoon
(D) In the morning
(E) At night

17. What does Tiger do when he hears a noise?

(A) He runs away
(B) He walks with stiff legs
(C) He hides under the bushes
(D) He walks on tiptoe
(E) He pretends he doesn't hear it

18. Who is afraid to come into his yard?

(A) All the other cats
(B) The dog next door
(C) The people who live in the house
(D) The mailman
(E) Most of the birds

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

The model number of your radio is A-707. Weak sound may indicate weak batteries. Replace with fresh batteries. Failure of the radio to operate may indicate a loose connection. All connections should be checked. If the radio still does not work properly, take it to our service department, 17-B West 17th Street.

19. What is the model number of the radio?

(A) A-707
(B) 17-B
(C) W-17
(D) B-17
(E) AB-707

20. What should be done if the sound is weak?

(A) Use weak batteries
(B) Send the model number to the service department
(C) Replace the present batteries with fresh batteries
(D) Check all the connections
(E) Replace the connections

21. What is the address of the service department?

(A) 17-A West 17th Street
(B) 17-B West 17th Street
(C) 17-A West 7th Street
(D) A-707 West 57th Street
(E) 17-B West 57th Street

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

Sara hated big dinners. There were so many dishes to wash afterwards, and no one ever thought to thank her for doing them. And people always stayed so late after a big dinner. Sometimes it was midnight before she could begin to clean up.

22. Why did Sara hate big dinners?

(A) Because she always ate too much
(B) Because people were so noisy
(C) Because there were so many dishes to wash
(D) Because she was never invited
(E) Because they were so expensive

23. How often did people remember to thank Sara?

(A) Sometimes
(B) Always
(C) Never
(D) Once
(E) Twice

24. How late did it sometimes get before Sara could clean up?

(A) Noon
(B) Morning
(C) Afternoon
(D) Midnight
(E) Evening

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

"I know you are in there," said the sheriff. "You have five seconds to come out."

"Come get me!" shouted the robber from inside the house.

The sheriff began to count. "One. Two. Three." Suddenly, the robber walked out with his hands up.

28. Where was the robber?

(A) Inside the house
(B) By the river
(C) In the bushes
(D) On his horse
(E) In the barn

29. How long did the sheriff give him to come out?

(A) Five seconds
(B) One minute
(C) Five minutes
(D) Ten minutes
(E) An hour

30. What did the robber do?

(A) He ran out shooting both guns
(B) He tried to escape and was shot down
(C) He walked out with his hands up
(D) He sneaked out and got away
(E) He didn't come out, so the sheriff had to go in and get him

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

The cat brushed against the old man. He did not move. He only stood, staring up into the window of the house. The party inside looked warm and friendly, but no one noticed him. The old man walked sadly on, followed by the cat.

25. What kind of animal was with the old man?

(A) Mouse
(B) Dog
(C) Horse
(D) Cat
(E) Bird

26. What was inside the house?

(A) A party
(B) Some dogs
(C) An old lady
(D) A meeting
(E) A salesman

27. The man is described as being

(A) Old
(B) Young
(C) Thin
(D) Fat
(E) Small

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

His cigarette went out. His pen dropped from his hand. His head began to nod. He was, all at once, asleep. Everyone in the room laughed, for he had come to work only five minutes ago.

31. What dropped from his hand?

(A) A pen
(B) A pencil
(C) A piece of paper
(D) A telephone
(E) A book

32. What was he doing after his head began to nod?

(A) Talking
(B) Sleeping
(C) Crying
(D) Smoking
(E) Leaving

33. When had he come to work?

(A) Half an hour ago
(B) Three hours ago
(C) Yesterday
(D) Five minutes ago
(E) Forty minutes ago

DO NOT TURN THE PAGE UNTIL YOU ARE TOLD TO DO SO.

APPENDIX V

FIVE ITEMS USED IN WRITING TEST

1. Turn left at the next corner.
2. School will be closed tomorrow because of heavy snow.
3. Send today for your free copy of this book.
4. If you need a doctor, call this number right away.
5. Drop a dime in the slot and turn the handle to the left.

APPENDIX VI

BASIC SKILLS SURVEY

READING AND WRITING

— MANUAL FOR EXAMINERS —

© Copyright 1966
by
Educational Testing Service
Princeton, N.J. Berkeley, Calif.
All rights reserved

BRIEF TEST OF LITERACY

MANUAL FOR EXAMINERS

INTRODUCTION

The Brief Test of Literacy was intended to provide a sound basis for classifying subjects as "literate" or "illiterate" within a very short time limit. There are two tests—one of reading and one of writing. The reading test contains seven brief paragraphs, each accompanied by three questions, for a total of twenty-one questions; the writing test consists of five sentences totaling forty-seven words.

The tests and testing procedures were designed to provide the maximum information for the simple categorical decision, "literate" or "illiterate." For both reading and writing, literacy was defined as approximately that level of function which is attained by the average student at the beginning of the fourth grade. Since the nature of the decision is essentially "either - or," a cutting score technique is used: all persons above a certain test score are classed as literate, all persons below the score are classed illiterate. The cutting score in turn provides the basis for the very brief testing times which are possible with this instrument, for the testing need only be continued until this score is achieved. That is, it is sufficient to be able to know that the subject is above the cutting score (hence "literate" by definition); how far above is not important. Indeed, the instrument is not well suited for differentiating among persons who are not near the cutting score. It tends to bunch such people into a single score category, since it has been specially built to provide its maximum of information at and near the cutting score. In achieving this maximum, information about differences at other levels is necessarily lost.

Required Materials

An administration requires:

(1) stopwatch
(2) pencils (with erasers)
(3) reading test booklet
(4) answer sheets
(5) manual for examiners

ADMINISTERING THE READING TEST

Procedures

Seat the subject at a desk or table, provide him with a pencil, answer sheet and booklet, and have him write his name in the space provided. Then say:

> This is a brief test of reading and writing. It will last about ten minutes. Read the instructions on the cover silently to yourself while I read them aloud to you.

Read as follows:

> On each page in this booklet there is a short paragraph which is followed by three questions. Below each question are five statements, only one of which makes a good and sensible answer. You should find this statement, and mark your answer by circling the letter on the answer sheet which corresponds to the statement you select.
>
> You must work as quickly as you can, for you will be allowed only one minute to work on each paragraph. Because the time is so short, you may not finish all of the questions. If you do finish a page before time is up, tell me and you will be allowed to go on to the next page.

After reading the instructions, ask if there are any questions. The typical subject will NOT have any questions; those who do will frequently merely require repetition of the appropriate part of the instructions. The following replies are suggested for possible questions in two areas:

Erasing

Questions: Can I change my answer?
Is it o.k. to erase?
Is it o.k. to cross out my first answer?
Reply: Yes, but work as quickly as you can.

Guessing

Questions: Is it o.k. to guess?
Do you count off for guessing?
Can I guess?

Reply: We are subtracting a penalty for each wrong answer, so wild guessing is unlikely to improve your score, and it may lower it. However, if you can eliminate one or more of the wrong answers, it is probably to your advantage to guess.

When the subject is ready, read the following, pointing to the appropriate section of the answer sheet:

> Read the paragraph and then answer the questions by circling the appropriate letters (point to 01, 02, 03 on the answer sheet). There is only one correct answer for each question. Tell me when you have finished with the paragraph.
>
> Ready? Begin work.

Begin timing. At the end of one minute, say:

> Stop working. The time is up. Do you have any questions?

If the subject completes the sample page in less than a minute, say:

> Finished? Fine. Do you have any questions?

Few questions will be asked. Some may inquire about guessing or erasing, as described above; a few may wonder if the paragraphs in the test are any longer than the sample paragraph. A simple reply is:

> The paragraphs differ in length from page to page, but they are all about as long as this sample.

When all is ready, say:

> Now we will begin the test. Remember, if you finish a page before time is called, tell me that you are finished. Do not turn to the next page until you are told to do so.
>
> Ready? Turn over to page one and begin working.

For each page, begin timing when the pages lie flat. Some subjects will smooth the booklet; others arrange their answer sheet; they vary in the way they spend the first few seconds. Therefore, there is a need for a fixed starting point, and this is when the pages lie flat. Do not worry if individual subjects seem to take too long before beginning work; the time allotted is really quite generous and any capable reader has sufficient time to demonstrate his ability.

The remaining work of giving the test is repetitive. If the subject indicates that he is finished, say:

> Finished? Turn over to page —— and begin work.

If the subject does not finish in one minute, say:

> Stop. Turn over to page ——and begin work.

Always state the page number which the subject should be working on, in order to avoid confusion.

Substitute Paragraphs

The administration of a rapidly-paced examination often leads to errors in timing, etc. In this examination, the time is so brief that a sneeze, a broken pencil, or other inadvertent interruption may cast doubt on the performance on a given paragraph. For this reason, alternate passages are provided on pages 8-11 of the test booklet. If one of the initial seven passages must be replaced, it is suggested that it be done according to the following program:

For Passage on Page	Use Passage on Page
1	8
2	9
3	11
4	10
5	9
6	9
7	9

The same cutting score of 11 may be used in each case. This procedure assumes an equivalence among passages that is not rigorously true. However, it would seem to be superior to the use of examiner judgment in effecting remedies for deviant records, for such judgments are characteristically unreliable.

Use of the Reading Test Cutting Score

This test is scored by giving 1 point for a right answer, 0 for an omit, and -¼ for a wrong answer. The complete reading test consists of seven passages, with a total of twenty-one questions. Because of the penalty for wrong answers, the scores could range from -5¼ (all wrong) to 21 (all right). In practice, however, all that we are interested in knowing is whether or not the subject gets a "formula score" (R-¼W) greater than 10.5. If he does, he passes and is classed "literate"; if his score is 10.5 or less, he fails and is "illiterate" in terms of this test. The cutting score was selected on the basis of the statistical information concerning the test.

Obviously, if the subject completes the first four paragraphs and gets all questions correct, he has a score of 12 and is "literate." There is no need to give additional questions. Similarly, if he gets 11 right and 1 wrong, he will pass. Almost all capable readers will answer the 12 questions correctly, and in much less time than the four minutes allotted. Hence, the use of the cutting score can reduce the average testing time for reading, including instructions, to under five minutes.

To use the cutting score, the examiner must be in position to observe the subject's work unobtrusively. In effect, he scores the answer sheet as the subject works. This is typically a simple operation and can be deferred until the fourth paragraph is begun. The scoring key is provided on page 6 of this manual.

Because the cutting score is between 10 and 11, it is possible to accept the decision "literate" before all questions on the fourth paragraph are completed. It is also possible to accept the other decision, "illiterate," before the fourth page is completed. (In fact, the decision "illiterate" may be reached at the conclusion of the first three passages, if all of the nine answers to these passages are wrong, for even if the subject answered the remaining twelve questions correctly, he would fail to achieve a score greater than the cutting score.) It is recommended, however, that full seven-passage records be obtained for all subjects excepting only those who have 11 or 12 right answers on the first four passages.

This recommendation means that even subjects who pass the cutting score in the course of their work on the fifth or sixth passage, should be continued for the full seven passages. It awards a premium, in a sense, to the perfect or near-perfect performance on the early paragraphs. Subjects who attain these excellent records may be presumed to be so capable that near-perfect performance on the remaining questions may be granted.

To summarize: the cutting score is between a formula score of 10.5 and one of 10.75; at 10.5 or less, the subject "fails" and is "illiterate," at 10.75 or greater, he "passes" and is "literate." Subjects will achieve the cutting score, or demonstrate an inability to achieve it at varying points in their work. It is recommended, however, that all subjects complete all seven passages excepting only those subjects who get 11 or 12 right answers on the first four passages. The time saving of the cutting score will be realized for a very large percentage of the prospective group, ages 12-17. Approximately 95% of this group may be anticipated to answer the twelve simple questions correctly and in a few short minutes. For the remainder of the group, the need for a complete record is more crucial and the attempt to save time by shortening the record is not worthwhile.

SCORING INFORMATION

Answer Keys

Sample Questions

01	A
02	A
03	E

Test Questions (Pages 1-7)

Page 1	Question	1	C
		2	A
		3	B
Page 2	Question	4	B
		5	C
		6	D
Page 3	Question	7	D
		8	A
		9	B
Page 4	Question	10	C
		11	D
		12	E
Page 5	Question	13	E
		14	A
		15	C
Page 6	Question	16	E
		17	B
		18	A
Page 7	Question	19	A
		20	C
		21	B

Supplementary Questions (Pages 8-11)

Page 8	Question	22	C
		23	C
		24	D
Page 9	Question	25	D
		26	A
		27	A
Page 10	Question	28	A
		29	A
		30	C
Page 11	Question	31	A
		32	B
		33	D

Decision Chart

After four passages:

Any subject having 11 or 12 right answers is classed "literate" and testing is discontinued.

After seven passages:

Any subject having 13 or more right answers is classed "literate."

Any subject having 12 or more right answers and 5 or fewer wrong answers is classed "literate."

Any subject having 11 right answers and only 1 or 0 wrong answers is classed "literate."

All other subjects are classed "illiterate."

ADMINISTERING THE WRITING TEST

Procedures

After the reading test is completed, say:

> That's the end of the reading test. The next test is the writing test. Turn over your answer sheet.

When the subject is ready, say the following, pointing to the three lines of the first answer space at the appropriate time:

> Listen carefully. I am going to read a sentence to you and I want you to write it in the space provided after I have read it twice. Use as much space as you need, and tell me if you want the sentence repeated. You have one minute.
>
> Do you have any questions?

Most questions seem to be quasi-questions which repeat the instructions in different wording and merely require some simple confirmation.

Example: "I write down what you say?"
Reply: "Yes."

Some subjects may ask: "Do you count off for poor spelling?"

A suggested reply would be: "Yes, spelling does count, but just do the best you can."

The questions which were mentioned earlier in connection with the reading test, concerning erasing, crossing out, etc., may also be asked at the beginning of the writing test. Refer to the earlier discussion for the suggested replies. Still another question may concern the possibility of breaking the pencil. If this is asked, say: "If you break your pencil, I will give you another."

When all is ready, read the sentence twice at a moderate rate. As you finish, say:

> Begin writing.

As you say "Begin writing," you should begin timing. Allow one minute and then say:

> Stop. Listen carefully and I will read sentence—— (the next sentence).

If the subject finishes before time is up, say:

> Finished? I will read sentence——.

In each case, say "Begin writing" as the signal to begin.

The sentences are:

1. Turn left at the next corner.
2. School will be closed tomorrow because of heavy snow.
3. Send today for your free copy of this book.
4. If you need a doctor, call this number right away.
5. Drop a dime in the slot and turn the handle to the left.

Some subjects will ask to have the sentence repeated. Others may have an obvious difficulty but hesitate to ask. The examiner should watch carefully and repeat the sentence on his own initiative if the subject appears to need it. This is not a memory test. No real harm can come from repetition. The average subject, of course, has no trouble retaining the sentence and would find further repetition an interruption.

In general, the writing test can be completed regardless of interruptions or breaking of pencils, etc., for if the examiner wishes he can always instruct the subject to begin over again and time him from the new start. That is, since the test is not one of memory, practice makes little difference, and a broken pencil or a fit of coughing or other interruption can be coped with by starting over again. If needed, the margins of the answer sheet will provide the space for a second attempt on interrupted questions.

SCORING INFORMATION

The scoring of constructed responses always poses difficulties, largely because of the variety of deviations from the norm which occur. Even in the simple task used in this test, the poorest writers will produce quite complex responses, difficult to evaluate. To reduce the problems and to secure reliability, the score for the writing test is based simply on the number of words correctly spelled and on the correctness of their order. For example, the sentence

If yu need a dokter, call this nmbr rite away.

receives a score of 6, 1 point for each correctly spelled (underlined) word. No credit is given for mis-

spellings, even when the approximations are as phonetically acceptable as the words "yu," "dokter," and "rite."

An immediate problem concerns the legibility of the handwriting. Inevitably, examiners will differ as to what the subject actually wrote. In general, the guiding principle should be to give the subject the benefit of the doubt on any given letter. That is, if the response to a word is so poorly written as to be meaningless, no credit for that word is given, but if a letter is unclear, do not penalize. For example, if it is uncertain whether the subject really wrote "e" for "o" in "doctor," give the subject the benefit of the doubt and 1 credit for the word. To repeat: If a single letter is ambiguous, assume that it is correct; if whole words are illegible, do not give credit.

The order of the words is important. The subject does not get credit even for a correctly spelled word if this word is out of place. For instance, if the example above had been written

> If yu need to call a dokter, use this nmbr rite away.

the score for this would be 5. The subject would not receive credit for call, which is out of place. This second example also provides an instance of the introduction of new words into the response, for "use" and "to" do not appear in the original sentence. No credit is lost for such introductions, which occur chiefly in the records of borderline subjects. The problem of determining the correctness of order is more difficult than is apparent at first glance. For example, one subject responded to sentence 1:

> Turn next at the left corner.

To cope with the diversity of possible subject responses, the scorer writes above each correctly spelled word the number which indicates the order in which it appears in the sentence as dictated. For example:

1 ⑤ 3 ④ 2 6
Turn next at the left corner.

This receives a score of 4, according to the following procedure: No word is scored if its number is greater than the number of the word immediately on its right. In the example above, "next" and "the" would not be scored, for "5" is greater than "3" and "4" is greater than "2." Because this procedure is basically mathematical and mechanical, it will not exclude the same words as would a judge. In the foregoing example a judge would rule out "next" and "left," rather than "next" and "the." However, the same score is arrived at both by judging and by applying the rule: four words are given credit. In problems of more complex reorderings, the merit of the mechanical approach will be apparent, for it is quite simple and reliable. A device for tallying the eliminated words is simply to draw a circle around the number above them.

One difficulty arises from the tendency of borderline subjects to repeat words. Thus, one candidate wrote

> Send today for this free book today.

This would be scored as 5:

1 2 3 ⑧ 5 9 [2]
Send today for this free book today.

There are two instances of the word today. Whenever this occurs, if you give credit for the first such word, by the basic rule, draw a square around the number of the second such word and omit it from further consideration in the scoring. By the normal application of the basic rule, the word "book" should not be scored, for its number, 9, is greater than the number of the next word. However, having credited the first "today," the second is deleted and does not affect the value of "book."

Some Examples of the Scoring

The following sentences were actually encountered in the testing:

2 3 ⑦ 4 9
Example 1 Sent today for fee cpy of your fee book.

The score is 4. Do not credit "of" because the number of the word to the right is less. Note that while actually "your" is more properly the misplaced word, the rule accounts for the inversion by excluding "of." The net effect is the same.

1 2 4 7 8 [7] 9
Example 2 Send today fou your charpy of this of book.

The score is 6. When the first "of" is scored, note its position number, 7, and draw a square around the second 7. This permits the word "this" to be scored when it is encountered later, for the next correctly spelled word is now "book" with a position number greater than that of "this."

1 2 3 4 5 6 8
Example 3 if you need a doctor call these number
10
write away.

The score is 8. Notice that punctuation errors, such as the failure to begin the sentence with a capital letter, are not penalized.

Note: In sentence 5 of the test, the word "the" appears three times. Therefore, the rule for dealing with redundancy and misplacements cannot be applied to this word in this sentence. After assigning the number to each word, apply the basic rule.

4 7 8 9 11 12
Example 4 Dron in and turn the hand to the life.

The score is 6. The second appearance of "the" is indexed as 12, and is not considered to be a redundant expression of the earlier appearance in which the word was indexed 9.*

The following example was contrived to clarify and demonstrate the scoring.

Example 5 Send tuday for copy of yore free copy of this book.

Step 1: Underline all correctly spelled words:

Send tuday for copy of yore free copy of this book.

Step 2: For each underlined word, write the number which indicates its position in the original sentence.

```
  .1          3   6  7          5   6  7
Send tuday for copy of yore free copy of
  8   9
this book.
```

Step 3: Begin scoring, counting any word which has a number less than that of the next correctly spelled word on its right. If you credit a word which appears twice, draw a square around its second appearance. For example, the following sequence would be followed in the sample sentence above.

Underlined word	Number	Action
Send	1	give credit
for	3	give credit
copy	6	give credit; box redundant second "6"
of	7	no credit (next number, 5, is less than 7); do not box second "7"
free	5	give credit
copy	6	no credit, because boxed
of	7	give credit
this	8	give credit
book	9	give credit

Step 4: Total the number of credits to get score. Score would be 7.

Note that the rule is not harsh. The subject could score at most 9 credits. He is penalized only for the misspellings, the only effect of the rules about misplacement being to avoid an overcredit for redundant correctly spelled words. Note also that a word is not boxed in its second appearance if it is not credited in the first place.

Summary of Scoring Rules

(1) Score one point for each correctly spelled word if the positional number is not circled or boxed.
(2) Circle any positional number which is greater than the positional number which next appears on the right. Ignore a positional number which is boxed.
(3) Box any positional number which has appeared earlier with a word which was credited. Do not box a number if it was not credited in its earlier appearance, or if it is a second or third appearance of the word "the" in the fifth sentence of the test.
(4) Do not penalize for punctuation.
(5) The score is the sum of all of the credited words.

The Cutting Score for Writing

The cutting score for writing is set between 27 and 28. Accordingly, a subject getting 27 is classed "illiterate"; a subject getting 28 is classed "literate." While it is possible to attain the cutting score before the five sentences are completed, no decision based on shortened records, analogous to the four-passage decision for reading, is suggested, for the savings in time would be negligible and the complexity of the scoring process would place too great a burden on the examiner.

Relationships between Reading and Writing; Reading the Best Single Index

The Brief Test of Literacy produces two scores, each yielding a judgment "literate." Because these two scores are not perfectly correlated, some subjects may be judged "literate" by one test but not by the other. If it is necessary to determine the relationship between literacy and some other variable, the conflict in the status of these cases must be resolved. On the basis of the available data and logical considerations as to the nature of the abilities, it is recommended that in any such cases the decision reached by means of the reading test be considered final. Thus, in relating literacy to age, for example, a subject who was "illiterate" in the light of the reading test, but "literate" in terms of the writing test would be classed "illiterate" in assessing the relationship in question.

* The examiner must use his judgment in assigning positional numbers to the word "the" in the fifth sentence of the test, if there are fewer than three "the's" in the response. Here it seems clear that the first "the" in the original sentence was omitted by the subject and that the "the's" in this response should be assigned the positional numbers of 9 and 12 rather than 5 and 9.

☆ U. S. GOVERNMENT PRINTING OFFICE — 1970 — 388-949 P. O. 31